

**Nois suppr ssor for speech identification system - subtracts microphone signal data from loudspeaker signal data**

Veröffentlichungsnr. (Sek.) DE4106405  
Veröffentlichungsdatum : 1991-09-26  
Erfinder : FUJIMOTO JUNICHIROH (JP); ARIYOSHI TAKASHI (JP)  
Anmelder : RICOH KK (JP)  
Veröffentlichungsnummer :  DE4106405  
Aktenzeichen:  
(EPIDOS-INPADOC-normiert) DE19914106405 19910228  
Prioritätsaktenzeichen:  
(EPIDOS-INPADOC-normiert) JP19900074689 19900323; JP19900074690 19900323  
Klassifikationssymbol (IPC) : G10L3/00  
Klassifikationssymbol (EC) : G10L21/02A1  
Korrespondierende Patentschriften

**Bibliographische Daten**

The noise suppression device has an audio device coupled to a loudspeaker (2) contained in the same room as the microphone (1) for the speech signals. Two signal extraction stages remove data from the signals fed to the loudspeaker (2) and provided by the microphone (1). The two signal extraction stages (10,20) are coupled directly and via a multiplier (27) with a controlled transfer function to a noise subtraction stage (30).

The microphone (1) is housed in vehicle together with the loudspeaker (2), the transfer function being adjusted for the number of people in the vehicle.

USE/ADVANTAGE - E.g. for voice-operated features of car telephone, or for voice control of vehicle functions. For reliable speech recognition in noisy environment.

Daten aus der esp@cenet Datenbank -- I2

(19) BUNDESREPUBLIK  
DEUTSCHLANDDEUTSCHES  
PATENTAMT(12) Patentschrift  
(10) DE 41 06 405 C2(51) Int. Cl. 6:  
**G 10 L 3/00**  
H 04 B 15/00  
B 60 R 27/00

DE 41 06 405 C2

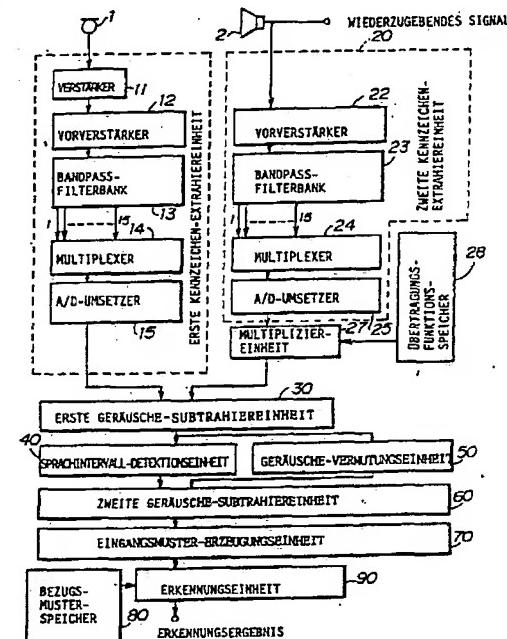
- (21) Aktenzeichen: P 41 06 405.4-53  
 (22) Anmeldetag: 28. 2. 91  
 (43) Offenlegungstag: 26. 9. 91  
 (45) Veröffentlichungstag der Patenterteilung: 29. 2. 96

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

(30) Unionspriorität: (32) (33) (31)  
23.03.90 JP 2-74689 23.03.90 JP 2-74690(73) Patentinhaber:  
Ricoh Co., Ltd., Tokio/Tokyo, JP(74) Vertreter:  
Schwabe, H., Dipl.-Ing.; Sandmair, K., Dipl.-Chem.  
Dr.jur. Dr.rer.nat.; Marx, L., Dipl.-Phys. Dr.rer.nat.,  
Pat.-Anwälte, 81677 München(72) Erfinder:  
Ariyoshi, Takashi, Yokohama, Kanagawa, JP;  
Fujimoto, Junichiroh, Yokohama, Kanagawa, JP(56) Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:  
DE 39 25 589 A1  
BOLL, S.F.: Suppression of Acoustic Noise in  
Speech Using Spectral Subtraction. In: IEEE  
Transactions on Acoustics, Speech, and Signal  
Processing, Vol. ASSP-27, Nr. 2, April 1979,  
S. 113-120;  
WIDROW, B. et al: Adaptive Noise Cancelling:  
Principles and Applications. In: Proceedings of the  
IEEE, Vol. 63, Nr. 12, Dez. 1975, S.1682-1716;  
FUJIMOTO, J., NAKATANI, T., YONEYAMA, M.:  
Speaker-Independent Word Recognition Using  
Fuzzy Pattern Matching. In: Fuzzy Sets and  
Systems 32 (1989) North-Holland, S. 181-191;

## (54) Geräuschunterdrückungseinrichtung für ein Spracherkennungssystem

(57) Geräuschunterdrückungseinrichtung für ein Spracherkennungssystem, in welchem eine Sprache eines Sprechers, welche einem Mikrofon (1) zugeführt worden ist, erkannt werden kann, wobei das Mikrofon in einem Raum vorgeschen ist, in dem sich ein Lautsprecher (2) befindet, welcher einen Klang abgibt, der einem wiederzugebenden Signal entspricht, welches von einer Audio-Einheit zugeführt wird, mit einer ersten Kennzeichnen-Extrahiereinheit (10), um erste Kennzeichendaten aus einem elektrischen Signal zu extrahieren, das von dem Mikrofon (1) abgegeben worden ist, wobei die ersten Kennzeichendaten ein Kennzeichen des elektrischen Signals darstellen; mit einer zweiten Kennzeichnen-Extrahiereinheit (20), um zweite Kennzeichendaten aus dem wiederzugebenden Signal zu extrahieren, welches dem Lautsprecher (2) zugeführt werden soll, wobei die zweiten Kennzeichendaten ein Kennzeichen des wiederzugebenden Signals darstellen, mit einer ersten Geräuschkomponenten erzeugenden Einheit, welche mit der zweiten Kennzeichnen-Extrahiereinheit (20) verbunden ist, um Geräuschedaten, welche auf den zweiten Kennzeichendaten von der zweiten Kennzeichnen-Extrahiereinheit (20) beruhen, zu erzeugen, wobei die Geräuschedaten einen Klang darstellen, welcher sich von dem Lautsprecher (2) durch den Raum zu dem Mikrofon (1) fortpflanzt, und mit einer ersten Geräusche-Subtrahiereinheit (30), welche mit der ersten Kennzeichnen-Extrahiereinheit (10) und der ersten Geräuschkomponenten erzeugenden Einheit verbunden ist, um die Geräuschedaten, welche von der ersten Geräuschkomponenten erzeugenden Einheit zugeführt worden sind, von den ersten Kennzeichendaten zu subtrahieren, die von der ersten Kennzeichnen-Extrahiereinheit (10) zugeführt worden sind, und um erste Sprachdaten abzugeben, so daß die ersten Sprachdaten verwendet werden, um die Sprache des Sprechers zu erkennen, dadurch gekennzeichnet, daß ein Speicher (28) zum Speichern von Übertragungskennedaten vorgesehen ist, welche einen Zustand darstellen, unter welchem sich ein Klang von dem Lautsprecher (2) zu dem Mikrofon (1) ...



## Beschreibung

Die Erfindung betrifft eine Geräuschunterdrückungseinrichtung für ein Spracherkennungssystem nach dem Oberbegriff des Anspruches 1.

Eine derartige Geräuschunterdrückungseinrichtung für ein Spracherkennungssystem ist aus der DE 39 25 589 A1 bekannt. Diese bekannte Geräuschunterdrückungseinrichtung umfaßt ein Mikrophon, welches zusammen mit wenigstens einem Lautsprecher in einem Raum aufgestellt ist, wobei der Lautsprecher einen Klang abgibt, der einem wiederzugebenden Signal entspricht, welches von einer Audio-Einheit zugeführt wurde. Es ist ferner eine erste Kennzeichen-Extrahiereinheit vorhanden, um erste Kennzeichendaten aus einem elektrischen Signal zu extrahieren, das von dem Mikrophon abgegeben worden ist, wobei die ersten Kennzeichendaten ein Kennzeichen des elektrischen Signals darstellen. Ferner ist eine zweite Kennzeichen-Extrahiereinheit vorhanden, um zweite Kennzeichendaten aus dem wiederzugebenden Signal zu extrahieren, welches dem Lautsprecher zugeführt werden soll, wobei die zweiten Kennzeichendaten ein Kennzeichen des wiederzugebenden Signals darstellen. Die zweite Kennzeichen-Extrahiereinheit ist mit einer ersten Geräuschkomponenten erzeugenden Einheit verbunden, um Geräuschdaten, welche auf den zweiten Kennzeichendaten von der zweiten Kennzeichen-Extrahiereinheit beruhen, zu erzeugen, wobei die Geräuschdaten einen Klang darstellen, welcher sich von dem Lautsprecher durch den Raum bis zu dem Mikrophon fortpflanzt. Mit der ersten Kennzeichen-Extrahiereinheit ist eine erste Geräusche-Subtrahiereinheit verbunden, welche auch mit der ersten Geräuschkomponenten erzeugenden Einheit verbunden ist, so daß Geräuschdaten, die von der ersten Geräuschkomponenten erzeugenden Einheit abgegeben werden, von den ersten Kennzeichendaten subtrahiert werden können, die von der ersten Kennzeichen-Extrahiereinheit zugeführt worden sind. Es lassen sich damit erste Sprachdaten erzeugen, die dazu verwendet werden können, um die Sprache des Sprechers zu erkennen.

Es ist auch das Eingeben von Informationen in eine Maschine mit Hilfe der Sprache bekannt geworden (siehe S. F. Boll: Suppression of Acoustic Noise in Speech Using Spectral Subtraction. In: IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-27, Nr. 2, April 1979, Seiten 113 bis 120, sowie B. Widrow et al.: Adaptive Noise Cancelling: Principles and Applications. In: Proceedings of the IEEE, Vol. 63, Nr. 12, Dez. 1975, Seiten 1692 bis 1716). Das heißt, es ist ein Spracherkennungssystem vorgeschlagen worden, welches bei einer Steuereinheit verwendet werden kann, um eine Audioeinheit, einen Luft-Conditioner bzw. eine Klimaanlage, ein Navigationssystem u. ä. zu steuern, welche in einem Fahrzeug vorgesehen sind. In dem Fahrzeug beeinflussen verschiedene Arten von Geräuschen im Fahrzeuginneren das Spracherkennungssystem, wenn das Fahrzeug fährt, und Klänge von in dem Fahrzeug vorgesehenen Audio-Einheiten wirken auf das Spracherkennungssystem, wie Geräusche im Fahrzeuginneren. Folglich muß in dem im Fahrzeug vorgesehenen Spracherkennungssystem eine Geräuschunterdrückungseinrichtung vorgesehen werden, um die verschiedenen Arten von Geräuschen im Fahrzeuginneren zu unterdrücken.

Beispiele von herkömmlicher Spracherkennung sind die "Spektrale Subtraktionsmethode von S. F. Boll u. a."

und eine "Adaptive Geräuschunterdrückungsmethode von B. Window u. a.".

In der "spektralen Subtraktionsmethode" wird ein Spektrum des Geräusches unter einer Bedingung abgetastet, bei welcher keine Sprache vorhanden ist, welche erkannt werden sollte, und dann wird das abgetastete Spektrum des Geräusches in einer Speichereinheit gespeichert. Das abgetastete Spektrum des Geräusches in der Speichereinheit wird fortlaufend aktualisiert. Wenn die Stimme eines Sprechers mit einem Mikrophon unter der Voraussetzung aufgenommen wird, daß es verschiedene Arten von Geräuschen gibt, wird das Spektrum des in der Speichereinheit gespeicherten Geräusches von einem akustischen Spektrum subtrahiert, welches einem akustischen Signal entspricht, das über das Mikrophon erhalten worden ist, so daß nur eine Stimmenspektrumskomponente, welche der Stimme des Sprechers entspricht, erhalten wird.

In der vorerwähnten "spektralen Subtraktionsmethode" können, da das Geräuschspektrum laufend aktualisiert wird, Hintergrundgeräusche von dem akustischen Signal subtrahiert werden. Es ist jedoch schwierig, ein unregelmäßiges Geräusch, welches bezüglich der Zeit in unregelmäßiger Weise erzeugt wird, von dem akustischen Signal zu subtrahieren. Das unregelmäßige Geräusch enthält den Klang, welcher von der Audio-Einheit, wie einem Radiogerät und einem Stereo-Bandrecorder, abgegeben worden ist.

Bei der "adaptiven Geräuschunterdrückungsmethode", bei welcher beispielsweise zwei Mikrophone verwendet werden, ist ein erstes Mikrophon nahe bei dem Sprecher und ein zweites Mikrophon weit von dem Sprecher entfernt vorgesehen. Die Stimme des Sprechers wird hauptsächlich von dem ersten Mikrophon aufgenommen. Verschiedene Arten von Geräuschen gelangen in die beiden Mikrophone bzw. werden von diesen aufgenommen. In diesem Fall wird ein Geräuschsignal, welches über das zweite Mikrophon erhalten worden ist, in einem adaptiven Filter verarbeitet, in welchem dessen Filtercharakteristik gesteuert werden kann. Das von dem adaptiven Filter abgegebene Signal wird von dem akustischen Signal subtrahiert, welches über das erste Mikrophon erhalten worden ist, so daß ein Stimmensignal, welches der Stimme des Sprechers entspricht, erhalten wird. Die Filtercharakteristik des vorerwähnten adaptiven Filters wird so gesteuert, daß der Fehler ( $\epsilon$ ) eines Ausgangssignals minimiert wird.

Bei der vorstehend erwähnten "adaptiven Geräuschunterdrückungsmethode" mit zwei Mikrofonen ist ein Computer für eine aufwendige Berechnung erforderlich, um die Filtercharakteristik des adaptiven Filters zu steuern. Daher werden die Kosten der Geräuschunterdrückungseinrichtung höher. In einem Fall, bei welchem das Geräusch, welches beachtet werden sollte, auf den Klang von der Audio-Einheit beschränkt wird, ist die Geräuschunterdrückungseinrichtung, die nach der vorerwähnten "adaptiven Geräuschunterdrückungsmethode" arbeitet, für den praktischen Gebrauch nicht gut. Außerdem ist es dann, wenn eine Vielzahl von Geräuschen zusammentrifft, schwierig, die Filtercharakteristik des vorerwähnten adaptiven Filters zu steuern, so daß keine richtige Rauschunterdrückung durchgeführt werden kann.

Die der Erfindung zugrunde liegende Aufgabe besteht darin, eine Geräuschunterdrückungseinrichtung für ein Spracherkennungssystem der angegebenen Gattung zu schaffen, welches die Möglichkeit bietet, auf schnelle Weise an gegebene Raumverhältnisse ange-

paßt werden zu können, um aus einem zu erzeugenden akustischen Signal Geräusche wirksamer entfernen zu können.

Diese Aufgabe wird erfahrungsgemäß durch die im Kennzeichnungsteil des Anspruches 1 aufgeführten Merkmale gelöst.

Besonders vorteilhafte Ausgestaltungen und Weiterbildungen der Erfindung ergeben sich aus den Unteransprüchen.

Im folgenden wird die Erfindung anhand von bevorzugten Ausführungsbeispielen unter Hinweis auf die Zeichnungen näher erläutert. Es zeigt:

Fig. 1 ein Blockdiagramm eines Spracherkennungssystems mit einer Geräuschunterdrückungseinrichtung gemäß einer ersten Ausführungsform mit Merkmalen nach der Erfindung;

Fig. 2 ein Blockdiagramm einer Modifikation des in Fig. 1 dargestellten Spracherkennungssystems;

Fig. 3 ein Blockdiagramm eines Spracherkennungssystems mit einer Geräuschunterdrückungseinrichtung gemäß einer zweiten Ausführungsform mit Merkmalen nach der Erfindung;

Fig. 4 ein Blockdiagramm einer Modifikation des in Fig. 3 dargestellten Spracherkennungssystems;

Fig. 5 ein Diagramm, in welchem Detektoren zum Feststellen des Öffnens und Schließens von Fahrzeuggütern dargestellt sind; und

Fig. 6 ein Diagramm eines Detektors, um festzustellen, wie oft ein Fahrzeug-Fenster geöffnet wird.

Nunmehr wird eine erste Ausführungsform mit Merkmalen nach der Erfindung anhand von Fig. 1 beschrieben. Das in Fig. 1 dargestellte Spracherkennungssystem ist beispielsweise in einem Fahrzeug vorgesehen. In Fig. 1 sind in dem Fahrzeug ein Mikrofon 1 und ein Lautsprecher 2 vorgesehen. Ein Radiogerät, ein Stereo-Gerät, u. ä. welche in dem Fahrzeug vorgesehen sind, liefern wiederzugebende Signale an den Lautsprecher 2, und der Lautsprecher 2 setzt die wiederzugebenden Signale in Klänge um. Die Stimme eines Sprechers, die Klänge aus dem Lautsprecher 2 und verschiedene Geräusche im Fahrzeuginnenraum werden von dem Mikrofon 1 aufgenommen. Ein erste Kennzeichen-Extrahiereinheit 10 holt Kennzeichen aus dem akustischen Signal heraus, welche Geräusche enthalten, wobei die akustischen Signale von dem Mikrofon 1 aus der ersten Kennzeichen-Extrahiereinheit 10 zugeführt werden. Eine zweite Kennzeichen-Extrahiereinheit 20 holt Kennzeichen aus wiederzugebenden Signalen heraus, welche von jeder Audio-Einheit, wie dem Radiogerät und dem Stereo-Gerät geliefert werden. Eine erste Geräusche-Subtrahiereinheit 30 subtrahiert Geräusche, welche von der zweiten Kennzeichen-Extrahiereinheit 20 erhalten worden sind, von Audio-Daten, welche von der ersten Kennzeichen-Extrahiereinheit 10 erhalten worden sind. Eine Sprachintervall-Detektionseinheit 40 stellt Sprachintervalle, welche Intervalle für Sprachen oder Stimmen sind, in einer Datenreihe fest, welche von der ersten Geräusche-Subtrahiereinheit 30 zugeführt worden sind. Eine Geräusche-Vermutungseinheit 50 vermutet Nicht-Sprachen-Intervalle, welche Intervalle außer dem Sprachintervall sind, in der Datenreihe, welche von der ersten Geräusche-Subtrahiereinheit 30 geliefert werden. Eine zweite Geräusche-Subtrahiereinheit 60 subtrahiert die Geräuschkomponente, welche durch die Vermutungsoperation in der Geräusche-Vermutungseinheit 50 erhalten worden sind, von den Daten in den Sprachintervallen, welche durch die Sprachintervall-Detektionseinheit 40 festgestellt

worin sind. Eine Eingangsmuster-Erzeugungseinheit 70 erzeugt ein eingegebenes Sprachmuster der Stimme des Sprechers auf der Basis der Daten, welche von der zweiten Geräusche-Subtrahiereinheit 60 abgegeben worden sind. Eine Erkennungseinheit 90 vergleicht verschiedene Referenzmuster, welche in einem Referenzmusterspeicher 80 gespeichert sind und das eingegebene Sprachmuster, das von der Eingangsmuster-Erzeugungseinheit 70 erzeugt worden ist, und gibt ein Erkennungsergebnis ab, welches durch den vorstehenden Vergleich erhalten wird.

Nunmehr wird das in Fig. 1 dargestellte Spracherkennungssystem im einzelnen beschrieben. Die erste Kennzeichen-Extrahiereinheit 10, welche die Merkmale oder

Kennzeichen aus dem von dem Mikrofon zugeführten, akustischen Signal extrahiert, weist einen Verstärker 11, eine Vorverzerrungsschaltung 12, eine Bandpaß-Filterbank 13, einen Multiplexer 14 und einen A/D-Umsetzer 15 auf. Der Verstärker 11 hat eine vorherbestimmte Verstärkung und verstärkt das akustische, von dem Mikrofon 1 zugeführte Signal. Die Vorverzerrerschaltung 12 hebt Frequenzkomponenten in dem Hochfrequenzband des von dem Verstärker 11 zugeführten, akustischen Signals an. Die Bandpaß-Filterbank 13 hat fünfzehn Bandpaß-FILTER. Die verschiedenen Frequenzbänder der fünfzehn Bandpaß-FILTER unterscheiden sich voneinander. Fünfzehn Frequenzen, welche auf einer Achse von logarithmischen Koordinaten, welche von 250 Hz bis 6,5 KHz reichen, in regelmäßigen Intervallen angeordnet sind, sind Mittenfrequenzen der Frequenzbänder der fünfzehn Bandpaß-FILTER. Die Bandpaß-Filterbank 13 hat auch Gleichrichter und Tiefpaßfilter, welche jeweils einem der Frequenzbänder entsprechen. Mit Hilfe der Bandpaß-FILTER stellt die Bandpaß-Filterbank 13 Spektren des von dem Mikrofon 1 aufgenommenen Klangs fest. Der Multiplexer 14 wählt eines der Spektren in den fünfzehn Frequenzbändern aus, wobei die Spektren von der Bandpaß-Filterbank 13 zugeführt werden. Der A/D-Umsetzer 15 setzt ein Spektrum in jedem Frequenzband mit einer Abtastfrequenz von 10 ms in digitale Daten um. Daher wird in der ersten Kennzeichen-Extrahiereinheit 10 das Signal, welches den Verstärker 11, die Vorverzerrerschaltung 12, die Bandpaß-Filterbank 13, den Multiplexer 14 und den A/D-Umsetzer 15 passiert hat, in ein zeitspektrales Muster  $X(t,f)$  umgesetzt, das Geräusche enthält, wobei t eine Zeitkomponente und f eine Frequenzkomponente ist.

Die zweite Kennzeichen-Extrahiereinheit 20 extrahiert die Merkmale aus dem wiederzugebenden Signal, welches durch den Lautsprecher 2 in den Klang umgesetzt werden soll. Der Klang aus dem Lautsprecher 2 wird als Geräusch dem Mikrofon 1 zugeführt. Die zweite Kennzeichen-Extrahiereinheit 20 weist eine Vorverzerrerschaltung 22, eine Bandpaß-Filterbank 23, einen Multiplexer 24 und einen A/D-Umsetzer 25 auf. Die Vorverzerrerschaltung 22 hebt die Frequenzkomponenten in dem hochfrequenten Band des wiederzugebenden Signals an, das von den Audio-Einheiten zugeführt worden ist. Die Bandpaß-Filterbank 23 ermittelt Spektren des wiederzugebenden Signals in derselben Weise wie die Bandpaß-Filterbank 13 der ersten Kennzeichen-Extrahiereinheit 10. Der Multiplexer 24 wählt eines der von der Bandpaß-Filterbank 23 gelieferten Spektren aus. Der A/D-Umsetzer 25 setzt ein Spektrum in jedem Frequenzband, wobei das Spektrum von dem Multiplexer 24 zugeführt wird, in digitale Daten um. In der zweiten Kennzeichen-Extrahiereinheit 20 wird das

wiederzugebende Signal, welches die Vorverzerrerschaltung 12, die Bandpaß-Filterbank 23, den Multiplexer 24 und den A/D-Umsetzer 25 passiert hat, in ein zeitspektrales Muster  $N(t,f)$  des wiederzugebenden Signals umgesetzt, das dem Lautsprecher 2 zugeführt werden soll.

Eine Übertragungsfunktion  $H(f)$  zwischen dem Lautsprecher 2 und dem Mikrofon 1 ist vorher in einem Übertragungsfunktions-Speicher 28 gespeichert worden: Die Übertragungsfunktion  $H(f)$  zeigt eine Voraussetzung an, bei welcher der von dem Lautsprecher 2 abgegebene Klang sich zu dem Mikrofon 1 hin ausbreitet. Das heißt, die Übertragungsfunktion  $H(f)$  zeigt eine Beziehung zwischen einem von dem Lautsprecher 2 abgegebenen Klang und einem an dem Mikrofon 1 ankommenden Klang an und verändert sie auf der Basis eines Zustands des Raums, welcher den Lautsprecher 2 und das Mikrofon 1 umgibt. Die Übertragungsfunktion  $H(f)$  kann experimentell erhalten werden. Wenn beispielsweise ein pulsförmiges Signal durch den Lautsprecher 2 in einen Klang umgewandelt wird, wird ein durch das Mikrofon 1 erhaltener Impuls-Frequenzbereich entsprechend einer Fourier-Transformation verarbeitet. Hierdurch wird die Übertragungsfunktion  $H(f)$  zwischen dem Lautsprecher 2 und dem Mikrofon 1 erhalten. Die Übertragungsfunktion  $H(f)$  zwischen dem Lautsprecher 2 und dem Mikrofon 1 kann auch auf der Basis eines Frequenzbereichs des Mikrofons 1 erhalten werden, wenn der Lautsprecher 2 ein weißes Rauschen oder Signale mit einer Frequenzabtastung wie-  
dergibt.

In einer Multipliziereinheit 27 wird das zeitspektrale Muster  $N(t,f)$  von dem A/D-Umsetzer 25 und die aus dem Speicher 28 gelesene Übertragungsfunktion  $H(f)$  eingegeben, und die Multipliziereinheit (27) berechnet ein erstes zeitspektrale Geräuschmuster  $N_1(t,f)$  entsprechend der folgenden Formel:

$$N_1(t,f) = N(t,f) \times H(f).$$

Das erste zeitspektrale Geräuschmuster  $N_1(t,f)$  entspricht einem zeitspektralen Muster eines Klangs, welcher von dem Mikrofon 1 aufgenommen wird, wenn der Lautsprecher 2 einen Klang mit dem zeitspektralen Muster  $N(t,f)$  abgibt.

Die erste Geräusche-Subtrahiereinheit 30 erhält das von der ersten Kennzeichen-Extrahiereinheit 10 abgegebene zeitspektrale Muster  $X(t,f)$  und das erste, von der zweiten Kennzeichen-Extrahiereinheit 20 abgegebene, zeitspektrale Geräuschmuster  $N_1(t,f)$  und subtrahiert das erste zeitspektrale Geräuschmuster  $N_1(t,f)$  von dem zeitspektralen Muster  $X(t,f)$ . Daher erzeugt die erste Geräusche-Subtrahiereinheit 30 ein zeitspektrales Muster  $X_1(t,f)$  entsprechend der nachstehenden Formel

$$X_1(t,f) = X(t,f) - N_1(t,f).$$

Wie vorstehend bereits beschrieben worden ist, entspricht das zeitspektrale Muster  $X_1(t,f)$  einem Signal, welches erhalten wird, indem das Geräusch, welches auf dem wiederzugebenden Signal basiert, welches durch den Lautsprecher 2 in den Klang umgewandelt werden soll, von dem akustischen Signal subtrahiert wird, das von dem Mikrofon 1 erhalten wird. Die Sprachintervall-Detektionseinheit 40 ermittelt die Sprachintervalle auf der Basis der zeitspektralen Muster  $X_1(t,f)$ , welche von der ersten Geräusche-Subtrahiereinheit abgegeben worden sind. Die Sprachintervall-Detektionseinheit 40

ermittelt die Sprachintervalle wie folgt:

In jedem Rahmen für eine Abtastzeit (beispielsweise 10 ms) werden die zeitspektralen Muster  $X_1(t,f)$  in allen (fünfzehn) Frequenzbändern so addiert, daß die folgende Gesamtsumme erhalten wird:

$$\sum_{f=1}^{15} X_1(f,t)$$

Wenn die Gesamtsumme größer als ein vorherbestimmter Schwellenwert ist, wird der Rahmen als das Sprachintervall festgelegt.

Die Geräusche-Vermutungseinheit 50 nimmt fortlaufend das Mittel der zeitspektralen Muster  $X_{1n}(t,f)$ , welche von der ersten Geräusche-Subtrahiereinheit 30 abgegeben werden, in Intervallen außer dem Sprachintervall, das von der Sprachintervall-Detektionseinheit 40 ermittelt worden ist. In den vorstehend angeführten Mustern  $X_{1n}(t,f)$  sind mit n die Intervalle außer den Sprachintervallen (d. h. keine Sprache enthaltende Intervalle bzw. Nicht-Sprachen-Intervalle) bezeichnet.  
20 Die Geräusche-Vermutungseinheit 50 gibt das Mittel der zeitspektralen Muster  $X_{1n}(t,f)$  als ein zweites zeitspektrale Geräuschmuster  $N_2(t,f)$  ab. Das zweite zeitspektrale Geräuschmuster  $N_2(t,f)$  entspricht inneren Geräuschen und nicht dem von dem Lautsprecher 2  
25 abgegebenen Klang. Das heißt, das zweite zeitspektrale Geräuschmuster  $N_2(t,f)$  entspricht einem Motorgeräusch, einem Straßenlärm u. ä., welches hauptsächlich regelmäßige Geräusche sind, in welchen sich über einen entsprechenden Zeitabschnitt bezüglich der Zeit  
30 nicht viel ändert. Die zweite Geräusche-Subtrahiereinheit 60 subtrahiert das zweite zeitspektrale Geräuschmuster  $N_2(t,f)$  von den zeitspektralen Mustern  $X_{1s}(t,f)$ , welche von der ersten Geräusche-Subtrahiereinheit 30 mittels der in der Sprachintervall-Detektionseinheit 40  
35 ermittelten Sprachintervalle abgegeben werden, so daß die zweite Geräusche-Subtrahiereinheit 60 die folgenden zeitspektralen Muster  $S(t,f)$  abgibt, welche der Stimme des Sprechers entsprechen:  
40

$$S(t,f) = X_{1s}(t,f) - N_2(t,f).$$

In dem vorstehenden Term  $X_{1s}(t,f)$  ist mit s das Zeitsprachintervall bezeichnet. Die Sprachintervall-Detektionseinheit 40, die Geräusche-Vermutungseinheit 50 und die zweite Geräusche-Subtrahiereinheit 60 arbeiten entsprechend der herkömmlichen "spektralen Subtraktionsmethode".

Die Eingangsmuster-Erzeugungseinheit 70 wandelt die Kennzeichen, welche durch die zeitspektralen Muster  $S(t,f)$  festgesetzt worden sind, welche von der zweiten Geräusche-Subtrahiereinheit 60 geliefert worden sind, in ein eingegebenes Sprachmuster entsprechend einem bekannten BTSP-(Binary Time Spectrum Pattern-)Spracherkennungsverfahren um. Das BTSP-Spracherkennungsverfahren ist beispielsweise in "Fuzzy Sets and Systems 32 (1989), Stn. 181 bis 191" (siehe 1989, Elsevier Science Publishers B.V., North-Holland) beschrieben. Der Referenzmusterspeicher 80 speichert die Referenzmuster. Jedes der Referenzmuster hat ein Format, das auf dem BTSP-Spracherkennungsverfahren basiert. Die Erkennungseinheit 90 gibt das eingegebene Sprachmuster von der Eingabemuster-Erzeugungseinheit 70 und die Referenzmuster von dem Referenzmu-

sterspeicher 80 ein und führt eine Erkennungsoperation auf der Basis des BTSP-Spracherkennungsverfahrens durch.

Die Sprachintervall-Detektionseinheit 40, die Geräusche-Vermutungseinheit 50, die zweite Geräusche-Subtrahiereinheit 60, die Eingangsmuster-Erzeugungseinheit 70 und die Erkennungseinheit 90 sind beispielsweise in einem Mikrocomputer ausgebildet. Außerdem können die Methode, die Sprachintervalle in der Sprachintervall-Detektionseinheit 40 zu detektieren, die Methode, das Geräusch in der zweiten Geräusche-Subtrahiereinheit 60 zu subtrahieren, das Verfahren, das eingegebene Sprachmuster in der Eingangsmuster-Erzeugungseinheit 70, das Format des in dem Referenzmusterspeicher 80 gespeicherten Musters und eine Erkennungsoperation in der Erkennungseinheit 90 zu erzeugen, in solche Verfahren abgeändert werden, welche auf dem Gebiet der Spracherkennung bekannt sind. Eine digitale Signalfunktionsschaltung, wie eine schnelle Fourier-Transformation (FFT), kann an die Stelle von jeder der Bandpaß-Filterbänke 13 und 23 gesetzt werden. Die ersten und zweiten Kennzeichen-Extrahiereinheiten 10 bzw. 20 haben die A/D-Umsetzer 15 und 25; die ersten und zweiten Kennzeichen-Extrahiereinheiten können jedoch auch statt dessen einen A/D-Umsetzer durch eine Time-Sharing-Steuerung gemeinsam benutzt.

Nunmehr wird eine Modifikation der ersten Ausführungsform anhand von Fig. 2 beschrieben. Fig. 2 zeigt das Spracherkennungssystem, um eine Sprache (Stimme) eines Sprechers unter einer Voraussetzung zu erkennen, bei welcher zwei Lautsprecher in dem Fahrzeug Klänge abgeben. In Fig. 2 sind die Teile, welche dieselben sind wie die in Fig. 1 dargestellten Teile mit denselben Bezeichnungen bezeichnet.

In Fig. 2 wird ein erstes wiederzugebendes Signal einem ersten Lautsprecher 2a und ein zweites wiederzugebendes Signal einem zweiten Lautsprecher 2b zugeführt. Das erste wiederzugebende Signal wird auch an eine zweite Kennzeichen-Extrahiereinheit 20a angelegt, und das zweite wiederzugebende Signal wird auch an eine andere, zweite Kennzeichen-Extrahiereinheit 20b angelegt. Das heißt, in diesem Spracherkennungssystem sind zwei zweite Kennzeichen-Extrahiereinheiten 20a und 20b vorgesehen. Eine zweite Kennzeichen-Extrahiereinheit 20a hat eine Vorverzerrerschaltung 22a, eine Bandpaß-Filterbank 23a, einen Multiplexer 24a und einen A/D-Umsetzer 25a in derselben Weise wie in Fig. 1 dargestellt ist. Die andere, zweite Kennzeichen-Extrahiereinheit 20b hat eine Vorverzerrerschaltung 22b, eine Bandpaß-Filterbank 23b, einen Multiplexer 24b und einen A/D-Umsetzer 25b in derselben Weise, wie in Fig. 1 dargestellt ist. Ein Speicher 28a speichert eine erste Übertragungsfunktion  $H_a(f)$  zwischen dem ersten Lautsprecher 2a und dem Mikrofon 1. Die erste Übertragungsfunktion  $H_a(f)$  zeigt einen Zustand an, bei welchem der von dem ersten Lautsprecher 2a abgegebene Klang sich zu dem Mikrofon 1 hin ausbreitet. Das heißt, die erste Übertragungsfunktion  $H_a(f)$  gibt eine Beziehung zwischen einem von dem ersten Lautsprecher 2a abgegebenen Klang und einem an dem Mikrofon 1 ankommenden Klang wieder und ändert sich aufgrund der Bedingungen in einem Raum, welcher den ersten Lautsprecher 2a und das Mikrofon 1 umgibt. Die erste Übertragungsfunktion  $H_a(f)$  kann auch experimentell in derselben Weise wie in dem anhand von Fig. 1 beschriebenen Fall erhalten werden. Ein Speicher 28b in der anderen zweiten Kennzeichen-Extrahierein-

heit 20b speichert eine zweite Übertragungsfunktion  $H_b(f)$  zwischen dem zweiten Lautsprecher 2b und dem Mikrofon 1. Die zweite Übertragungsfunktion  $H_b(f)$  gibt einen Zustand wieder, unter welchem sich der von dem zweiten Lautsprecher 2b abgegebene Klang zu dem Mikrofon 1 hin ausbreitet bzw. fortpflanzt. Das heißt, die zweite Übertragungsfunktion  $H_b(f)$  gibt eine Beziehung zwischen einem von dem zweiten Lautsprecher 2b abgegebenen Klang und einem an dem Mikrofon 1 ankommenden Klang wieder und ändert sich aufgrund des Zustandes bzw. der Voraussetzungen in einem Raum, welcher den zweiten Lautsprecher 2b und das Mikrofon 1 umgibt. Die zweite Übertragungsfunktion  $H_b(f)$  kann auch auf dieselbe Weise wie in dem unter Bezugnahme auf Fig. 1 beschriebenen Fall experimentell erhalten werden.

In der zweiten Kennzeichen-Extrahiereinheit 20a wird ein zeitspektrales Muster  $N_a(t,f)$  des ersten wiederzugebenden Signals erhalten, und dann multipliziert eine Multipliziereinheit 27a das zeitspektrale Muster  $N_a(t,f)$  und die erste Übertragungsfunktion  $H_a(f)$ , so daß dann die Multipliziereinheit 27a abgibt:  $N_a(t,f) \times H_a(f)$ . In der anderen, zweiten Kennzeichen-Extrahiereinheit 20b wird ein zeitspektrales Muster  $N_b(t,f)$  des zweiten wiederzugebenden Signals erhalten, und dann multipliziert eine Multipliziereinheit 27b die zeitspektren Muster  $N_b(t,f)$  und die zweite Übertragungsfunktion  $H_b(f)$ , so daß dann die Multipliziereinheit 27b abgibt:  $N_b(t,f) \times H_b(f)$ . Ein Addierer 29 erhält  $N_a(t,f) \times H_a(f)$  von der Multipliziereinheit 27a und  $N_b(t,f) \times H_b(f)$  von der Multipliziereinheit 27b und gibt das folgende, zeitspektrale Geräuszmuster  $N_1(t,f)$  ab:

$$N_1(t,f) = N_a(t,f) \times H_a(f) + N_b(t,f) \times H_b(f).$$

Der Spracherkennungsprozeß wird dann auf der Basis des zeitspektren Musters  $X(t,f)$ , das von der ersten Kennzeichen-Extrahiereinheit 10 abgegeben worden ist, auf der Basis des zeitspektren Geräuszmusters  $N_1(t,f)$ , das von dem Addierer 29 abgegeben worden ist, und der Referenzmuster aus dem Speicher 80 in derselben Weise wie bei dem vorstehend anhand von Fig. 1 beschriebenen Prozeß durchgeführt.

Das in Fig. 2 dargestellte Spracherkennungssystem wird in dem Fall angewendet, wenn zwei Lautsprecher 2a und 2b im Inneren des Fahrzeugs vorgesehen sind. Selbst wenn die Anzahl von im Inneren des Fahrzeugs vorgesehener Lautsprecher größer als zwei ist, kann jedoch das Spracherkennungssystem in derselben Weise wie in Fig. 2 dargestellt ist, realisiert werden. Das heißt, eine zweite Kennzeichen-Extrahiereinheit, eine Multipliziereinheit und einen Übertragungsfunktionspeicher, welcher jedem der Lautsprecher entspricht, sind vorgegeben und Ausgangssignale von den verschiedenen zweiten Kennzeichen-Extrahiereinheiten werden so addiert, daß das zeitspektrale Geräuszmuster  $N_1(t,f)$  erhalten wird.

Nunmehr wird eine zweite Ausführungsform mit Merkmalen nach der Erfindung anhand von Fig. 3 beschrieben. Die Übertragungsfunktion  $H(f)$  zwischen dem Lautsprecher 2 und dem Mikrofon 1 ändert sich aufgrund des Zustands bzw. der Voraussetzungen eines Raums, welcher den Lautsprecher 2 und das Mikrofon 1 umgibt. Menschliche Körper absorbieren eine akustische Welle, so daß sich beispielsweise in dem Fahrzeug die Übertragungsfunktion  $H(f)$  in Abhängigkeit von der Anzahl der Personen im Inneren des Fahrzeugs ändert. Die Übertragungsfunktion  $H(f)$  ändert sich auch in Zu-

sammenhang mit der Anzahl geöffneter Türen und/oder Fenster in dem Fahrzeug. Daher werden in der zweiten Ausführungsform eine Anzahl Übertragungsfunktionen in dem Übertragungsfunktions-Speicher gespeichert, wobei jede der Übertragungsfunktionen der Anzahl Personen oder der Anzahl geöffneter Türen und/oder Fenster in dem Fahrzeug entspricht.

In Fig. 3 sind die Teile, welche dieselben wie diejenigen in Fig. 1 sind, mit denselben Bezugszeichen bezeichnet. In Fig. 3 werden eine Anzahl Übertragungsfunktionen  $H_r(t,f)$  vorher in dem Übertragungsfunktions-Speicher 28 gespeichert. In den Funktionen  $H_r(t,f)$  ist mit  $r$  ein bestimmter Typ der Übertragungsfunktionen  $H(t,f)$  bezeichnet. Beispielsweise wird dann, wenn die Anzahl Personen im Inneren des Fahrzeugs gleich  $r$  ( $r = 1, 2, \dots$ ) ist, die Übertragungsfunktion mit  $H_r(f)$  bezeichnet. Jede der Übertragungsfunktionen  $H_r(f)$  wird experimentell unter der Voraussetzung erhalten, bei welcher  $r$  Personen in einem Fahrzeug sitzen. Beispielsweise wird jede der Übertragungsfunktionen  $H_r(f)$  auf der Basis des Impuls-Frequenzbereichs zwischen dem Lautsprecher 2 und dem Mikrophon 1 unter der Voraussetzung erhalten, daß  $r$  Personen in dem Fahrzeug sitzen.

Ein Wähltschalter (Wähleinheit) 28' ist zum Auswählen einer der Übertragungsfunktionen  $H_r(f)$  vorgesehen. Der Wähltschalter 28' hat eine Anzahl von Schaltelementen, welche jeweils der Anzahl Personen in dem Fahrzeug entsprechen. Wenn beispielsweise  $r$  Personen in dem Fahrzeug sitzen, kann der Fahrer des Fahrzeugs ein Schaltelement betätigen, das  $r$  Personen entspricht. Infolge der Betätigung des  $r$  Personen entsprechenden Schaltelements in dem Wähltschalter 28' wird die Übertragungsfunktion  $H_r(f)$  aus dem Speicher 28 ausgelesen und der Multipliziereinheit 27 zugeführt. Die Multipliziereinheit 27 berechnet dann das folgende zeitspektrale Geräuschmuster  $N_1(t,f)$ .

$$N_1(t,f) = N(t,f) \times H_r(f).$$

Folglich wird dieses zeitspektrale Geräuschmuster  $N_1(t,f)$  von der zweiten Kennzeichen-Extrahiereinheit 20 abgegeben.

Dann werden die Spracherkennungsprozesse auf der Basis des zeitspektralen Musters  $X(t,f)$ , das von der ersten Kennzeichen-Extrahiereinheit 10 abgegeben worden ist, auf der Basis des zeitspektralen Geräuschmusters  $N_1(t,f)$ , das von der zweiten Kennzeichen-Extrahiereinheit 20 abgegeben worden ist, und auf der Basis des von dem Speicher 80 abgegebenen Referenzmusters in derselben Weise wie die anhand von Fig. 1 beschriebenen Prozesse durchgeführt.

Bei der zweiten Ausführungsform kann die Übertragungsfunktion entsprechend der Anzahl Personen, welche in dem Fahrzeug sitzen, ausgewählt werden, so daß absolut sicher das Geräusch von dem Lautsprecher 2 von dem Geräusch an dem Mikrophon 1 bzw. das von dem Mikrophon 1 aufgenommen worden ist, subtrahiert wird, selbst wenn sich die Anzahl Personen im Inneren des Fahrzeugs ändert.

Nunmehr werden verschiedene Modifikationen der zweiten Ausführungsform anhand von Fig. 4 bis 6 beschrieben. In Fig. 4 ist ein Spracherkennungssystem dargestellt, bei welchem zwei Lautsprecher 2a und 2b im Inneren des Fahrzeugs vorgesehen sind. In Fig. 4 sind die Teile, welche dieselben sind wie die in Fig. 2 und 3 dargestellten Teile mit denselben Bezugszeichen bezeichnet. In diesem Fall hat das Spracherkennungssystem zwei zweite Kennzeichen-Extrahiereinheiten 20a

und 20b, welche den jeweiligen Lautsprechern 2a und 2b in derselben Weise wie diejenigen in Fig. 2 entsprechen. In jeder der zweiten Kennzeichen-Extrahiereinheiten 20a und 20b werden eine Anzahl Übertragungsfunktionen  $H_r(f)$  in dem jeweiligen Übertragungsfunktions-Speicher 28a und 28b in derselben Weise wie in dem anhand von Fig. 2 beschriebenen Spracherkennungssystem gespeichert. Dann wählt der Wähltschalter 28' eine der in dem jeweiligen Speicher 28a und 28b gespeicherten Übertragungsfunktionen in derselben Weise aus, wie in Fig. 3 dargestellt und anhand von Fig. 3 beschrieben ist.

Die Übertragungsfunktion ändert sich in Abhängigkeit von einem Zustand, ob nämlich jede der Türen und/oder Fenster des Fahrzeugs geöffnet oder geschlossen ist. In diesem Fall werden eine Anzahl Übertragungsfunktionen  $H_r(f)$  vorher in dem in Fig. 3 dargestellten Speicher 28 oder in jedem der in Fig. 4 dargestellten Speicher 28a und 28b gespeichert. Beispielsweise gibt es verschiedene Zustände, daß nämlich jede der Türen des Fahrzeugs entweder geöffnet oder geschlossen ist. Wenn das Fahrzeug V vier Türen D1 bis D4 hat, wie in Fig. 5 dargestellt ist, gibt es 16 ( $2^4$ ) Zustände, da jede der Türen D1 bis D4 entweder geöffnet oder geschlossen sein kann. Folglich entspricht jede der Übertragungsfunktionen  $H_r(f)$  dem  $r$ -ten Zustand der Türen ( $r=1, 2, \dots, 16$ ). Jede der Übertragungsfunktionen  $H_r(f)$  wird beispielweise auf der Basis des Impuls-Frequenzverlaufs zwischen jedem Lautsprecher und dem Mikrophon 1 in dem  $r$ -ten Zustand der Türen erhalten.

In diesem Fall sind in dem Fahrzeug V, wie in Fig. 5 dargestellt ist, vier Schalter SW1 bis SW4 vorgesehen, um festzustellen ob verschiedene Türen D1 bis D4 geöffnet sind oder nicht. Jeder der Schalter SW1 bis SW4 wird automatisch angeschaltet, wenn eine der entsprechenden Türen D1 bis D4 geschlossen ist. Ein-/Aus-Signale, welche von den Schaltern SW1 bis SW4 abgegeben werden, werden einer Steuereinheit 100 zugeführt. Die Steuereinheit 100 hat Zugriff zu dem Speicher 28 und wählt eine Übertragungsfunktion  $H_r(f)$  aus, welche dem  $r$ -ten Zustand der Türen entspricht, welcher durch die Ein-/Aus-Signale spezifiziert ist, welche durch die Schalter SW1 bis SW4 geliefert worden sind. Die ausgewählte Übertragungsfunktion  $H_r(f)$  wird an die Multipliziereinheit 27 angelegt, welche dann das folgende zeitspektrale Geräuschmuster  $N_1(t,f)$  in derselben Weise wie in Fig. 1 dargestellt ist, berechnet, nämlich:

$$N_1(t,f) = X(t,f) \times H_r(f).$$

Es kann auch eine Übertragungsfunktion zwischen dem Lautsprecher und dem Mikrophon entsprechend verschiedenen Zuständen bestimmt werden, je nachdem welches der Fenster des Fahrzeugs geöffnet oder geschlossen ist. Außerdem kann die Übertragungsfunktion zwischen dem Lautsprecher und dem Mikrophon in Übereinstimmung damit bestimmt werden, wie viele der verschiedenen Fenster, die in dem Fahrzeug vorgesehen sind, geöffnet sind. In diesem Fall ist eine Fühleinrichtung, um festzustellen, wie viele der Fenster jeweils geöffnet sind, in dem Spracherkennungssystem vorgesehen, wie in Fig. 6 dargestellt ist.

In Fig. 6 hat eine Fühleinrichtung 40a einen Schiebewiderstand R, welcher an einem Rahmen 41 jedes Fensters WD vorgesehen ist, eine Gleichstromquelle Vo und ein Ampermeter 42. Das Fenster WD kann in dem Rahmen auf- und abwärts bewegt werden. Der Wert des Schiebewiderstands R wird entsprechend der Größe

der Kante des Fensters WD geändert, die mit dem Schiebewiderstand R in Kontakt steht. Die Gleichstromquelle Vo, der Schiebewiderstand R und das Ampermeter 42 sind in Reihe geschaltet, so daß der Wert des Schiebwiderstand R über das Ampermeter 42 ermittelt wird. Das heißt, ein Ausgangssignal von dem Ampermeter 42 gibt wieder, wie weit jedes Fenster WD geöffnet ist. Das Ausgangssignal von dem Ampermeter 42 wird der Steuereinheit 100 zugeführt, welche wiederum die Übertragungsfunktion auswählt, welche einem Ausgangssignal entspricht, das von jeder Fühleinheit 40a geliefert worden ist, die an jedem der Fenster des Fahrzeugs vorgesehen ist.

Um zu verhindern, daß die Anzahl an Übertragungsfunktionen zunimmt, ist es wünschenswert, daß die Öffnungsweite jedes Fensters WD in zwei oder drei Stufen eingeteilt wird. Beispielsweise ist in der ersten Stufe jedes Fenster geschlossen oder befindet sich in einem Zustand zwischen dem geschlossenen Zustand und einem Zustand, in welchem es zu einem Drittel geöffnet ist. In der zweiten Stufe befindet sich jedes Fenster in dem Zustand, daß es zu einem Drittel geöffnet ist, oder weist einen Zustand zwischen dem Zustand, in welchem es zu einem Drittel geöffnet ist, und einem Zustand, in welchem es zu zwei Dritteln geöffnet ist. In einer dritten Stufe nimmt jedes Fenster einen Zustand ein, in welchem es zu zwei Dritteln geöffnet ist, oder es ist vollständig geöffnet oder nimmt einen Zustand zwischen dem Zustand, in welchem es zu zwei Dritteln geöffnet ist, und dem vollständig geöffneten Zustand ein. Wenn in diesem Fall vier Fenster in dem Fahrzeug vorgesehen sind, haben die Fenster 81 (3<sup>4</sup>) Zustände. Jede Übertragungsfunktion, welche jedem der 81 Zustände der Fenster entspricht, ist dann in dem Übertragungsfunktions-Speicher 28 vorgesehen.

Wieweit jedes Fenster geöffnet ist, kann auch mit Hilfe optischer Sensoren festgestellt werden. Außerdem kann die Übertragungsfunktion auf der Basis einer Kombination von zwei oder mehr der folgenden Faktoren bestimmt werden: die Anzahl Personen im Inneren des Fahrzeugs, die Anordnung der Personen im Inneren des Fahrzeugs, ob jede Tür geöffnet oder geschlossen ist, und wie weit jedes Fenster geöffnet ist.

Gemäß der Erfindung kann das Geräusch von dem Lautsprecher von dem Klang, welcher dem Mikrofon zugeführt bzw. von diesem aufgenommen wird, mit Hilfe der Übertragungskenndaten, wie der Übertragungsfunktion zwischen Lautsprecher und dem Mikrofon, subtrahiert werden. Somit kann ohne weiteres das Geräusch gegenüber dem akustischen Signal unterdrückt werden.

#### Patentansprüche

1. Geräuschunterdrückungseinrichtung für ein Spracherkennungssystem, in welchem eine Sprache eines Sprechers, welche einem Mikrofon (1) zugeführt worden ist, erkannt werden kann, wobei das Mikrofon in einem Raum vorgesehen ist, in dem sich ein Lautsprecher (2) befindet, welcher einen Klang abgibt, der einem wiederzugebenden Signal entspricht, welches von einer Audio-Einheit zugeführt wird, mit einer ersten Kennzeichen-Extrahiereinheit (10), um erste Kennzeichendaten aus einem elektrischen Signal zu extrahieren, das von dem Mikrofon (1) abgegeben worden ist, wobei die ersten Kennzeichendaten ein Kennzeichen des elektrischen Signals darstellen; mit einer zweiten

Kennzeichen-Extrahiereinheit (20), um zweite Kennzeichendaten aus dem wiederzugebenden Signal zu extrahieren, welches dem Lautsprecher (2) zugeführt werden soll, wobei die zweiten Kennzeichendaten ein Kennzeichen des wiederzugebenden Signals darstellen, mit einer ersten Geräuschkomponenten erzeugenden Einheit, welche mit der zweiten Kennzeichen-Extrahiereinheit (20) verbunden ist, um Geräuschedaten, welche auf den zweiten Kennzeichendaten von der zweiten Kennzeichen-Extrahiereinheit (20) beruhen, zu erzeugen, wobei die Geräuschedaten einen Klang darstellen, welcher sich von dem Lautsprecher (2) durch den Raum zu dem Mikrofon (1) fortpflanzt, und mit einer ersten Geräusche-Subtrahiereinheit (30), welche mit der ersten Kennzeichen-Extrahiereinheit (10) und der ersten Geräuschkomponenten erzeugenden Einheit verbunden ist, um die Geräuschedaten, welche von der ersten Geräuschkomponenten erzeugenden Einheit zugeführt worden sind, von den ersten Kennzeichendaten zu subtrahieren, die von der ersten Kennzeichen-Extrahiereinheit (10) zugeführt worden sind, und um erste Sprachdaten abzugeben, so daß die ersten Sprachdaten verwendet werden, um die Sprache des Sprechers zu erkennen, dadurch gekennzeichnet, daß ein Speicher (28) zum Speichern von Übertragungskenndaten vorgesehen ist, welche einen Zustand darstellen, unter welchem sich ein Klang von dem Lautsprecher (2) zu dem Mikrofon (1) ausbreitet, wobei die Übertragungskenndaten vorher in dem Raum gemessen worden sind, in welchem das Mikrofon (1) und der Lautsprecher (2) vorgesehen werden, und daß die erste Geräuschkomponenten erzeugende Einheit auch mit dem Speicher (28) verbunden ist, um Geräuschedaten, welche auf den zweiten Kennzeichen-Extrahiereinheit (20) und den Übertragungskenndaten von dem Speicher (28) beruhen, zu erzeugen.

2. Einrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die Übertragungskenndaten eine Übertragungsfunktion zwischen dem Lautsprecher (2) und dem Mikrofon (1) angeben, wobei die Übertragungsfunktion eine Beziehung zwischen einem Klang, welcher von dem Lautsprecher (2) abgegeben wird, und einem Klang darstellt, welcher von dem Lautsprecher (2) dem Mikrofon (1) zugeführt wird.

3. Einrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die erste Geräuschkomponenten erzeugende Einheit eine Multipliziereinheit (27) aufweist, welche die zweiten Kennzeichendaten und die Übertragungsfunktion miteinander multipliziert, wobei das Multiplizierergebnis in der Multipliziereinheit (27) als die Geräuschedaten verwendet werden.

4. Einrichtung nach Anspruch 1, dadurch gekennzeichnet, daß der Speicher (28) eine Anzahl Übertragungskenndaten speichert, welche jeweils einem Zustand des Raumes entsprechen, in welchem der Lautsprecher (2) und das Mikrofon (1) vorgesehen sind, und daß eine Wähleinheit (28') vorgesehen ist, um eine der Übertragungskenndaten aus dem Speicher (28) entsprechend einem Zustand des Raumes auszuwählen, in welchem der Lautsprecher (2) und das Mikrofon (1) vorhanden sind, so daß die ausgewählten Übertragungskenndaten der ersten Geräuschkomponenten erzeugenden Ein-

heit zugeführt werden.

5. Einrichtung nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß das Mikrophon (1) im Inneren eines Fahrzeugs vorgesehen ist, in welchem der Lautsprecher (2) vorgesehen ist, und daß jede der Übertragungskenndaten auf der Basis der Anzahl Personen im Inneren des Fahrzeugs bestimmt wird, und daß die Wähleinheit (28') die Übertragungskenndaten auswählt, welche der Anzahl Personen im Inneren des Fahrzeugs entsprechen. 10

6. Einrichtung nach Anspruch 5, dadurch gekennzeichnet, daß Übertragungskenndaten jeweils in Abhängigkeit davon bestimmt werden, ob jede Tür des Fahrzeugs geöffnet oder geschlossen ist, und daß die Wähleinheit (28') diejenigen Übertragungskenndaten auswählt, welche dem Zustand entsprechen, bei welchem die Türen des Fahrzeugs geöffnet oder geschlossen sind. 15

7. Einrichtung nach Anspruch 6, dadurch gekennzeichnet, daß die Wähleinheit (28') Fühleinheiten (SW1 bis SW4; 100) aufweist, um festzustellen, ob jede Tür des Fahrzeugs geöffnet ist oder nicht, so daß die jeweilige Übertragungscharakteristik auf der Basis eines Fühlergebnisses in der Fühleinheit (SW1 bis SW4; 100) gewählt wird. 20

8. Einrichtung nach Anspruch 4, dadurch gekennzeichnet, daß Übertragungskenndaten jeweils in Abhängigkeit davon bestimmt werden, wie weit jedes Fenster (WD) des Fahrzeugs geöffnet ist, und daß die Wähleinheit (28') die jeweiligen Übertragungskenndaten in Abhängigkeit davon auswählt, wie weit jedes Fenster des Fahrzeugs geöffnet ist. 30

9. Einrichtung nach Anspruch 8, dadurch gekennzeichnet, daß die Wähleinheit (28') eine Fühleinheit (40a) hat, um festzustellen, wie weit jedes Fenster des Fahrzeugs geöffnet ist, so daß eine Übertragungscharakteristik in Abhängigkeit davon gewählt wird, wie weit jedes Fenster geöffnet ist. 35

10. Einrichtung nach Anspruch 1, gekennzeichnet durch eine zweite Geräuschkomponenten erzeugende Einheit (40, 50), welche mit der ersten Geräusche-Subtrahiereinheit (30) verbunden ist, um eine Geräuschkomponente auf der Basis der ersten Sprachdaten zu erzeugen, und eine zweite Geräusche-Subtrahiereinheit (60), um die Geräuschkomponente, welche von der zweiten Geräuschkomponenten erzeugenden Einheit (40, 50) erzeugt worden ist, von den ersten Sprachdaten zu subtrahieren, welche von der ersten Geräusche-Subtrahiereinheit (30) zugeführt worden sind, wobei die zweite Geräusche-Subtrahiereinheit (60) zweite Sprachdaten abgibt, so daß die zweiten Sprachdaten zum Erkennen der Sprache/Stimme des Sprechers verwendet werden. 40 45 50 55

11. Einrichtung nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß in dem Raum mehrere Lautsprecher (2a, 2b) mit einem Mikrophon (1) kombiniert angeordnet sind und daß Speicher (28a, 28b) zum Speichern von Übertragungskenndaten ausgebildet sind, welche einen Zustand darstellen, bei welchem ein Klang sich von jedem der Lautsprecher (2a, 2b) zu dem Mikrophon (1) fortpflanzt, wobei die Übertragungskenndaten vorher in dem Raum gemessen worden sind, in welchem das Mikrophon (1) und die Lautsprecher (2a, 2b) angeordnet sind. 60 65

12. Einrichtung nach Anspruch 11, dadurch gekenn-

zeichnet, daß die Übertragungskenndaten eine Übertragungsfunktion zwischen jedem der Lautsprecher (2a, 2b) und dem Mikrophon (1) sind, wobei die Übertragungsfunktion eine Beziehung zwischen einem Klang, welcher von jedem der Lautsprecher (2a, 2b) abgegeben wird, und einem Klang darstellen, welcher von jedem der Lautsprecher (2a, 2b) dem Mikrophon (1) zugeführt wird.

13. Einrichtung nach Anspruch 3, dadurch gekennzeichnet, daß die Geräuschdaten erzeugende Einheit einen Addierer (29) aufweist, um Geräuschkomponenten, welche von der ersten Geräuschkomponenten erzeugenden Einheit zugeführt worden sind, zu addieren, wobei dann der Addierer (29) die Geräuschdaten abgibt.

14. Einrichtung nach Anspruch 11, gekennzeichnet durch eine zweite Geräuschkomponenten erzeugende Einheit (40, 50), welche mit der ersten Geräusche-Subtrahiereinheit (30) verbunden ist, um eine Geräuschkomponente auf der Basis der ersten Sprachdaten zu erzeugen, und eine zweite Geräusche-Subtrahiereinheit (60) zum Subtrahieren der Geräuschkomponente, welche von der zweiten Geräuschkomponenten erzeugenden Einheit (40, 50) erzeugt worden ist, von den ersten Sprachdaten zu subtrahieren, welche von der ersten Geräusche-Subtrahiereinheit (30) zugeführt worden sind, wobei die zweite Geräusche-Subtrahiereinheit (60) zweite Sprachdaten abgibt, so daß die zweiten Sprachdaten zum Erkennen der Sprache/Stimme des Sprechers verwendet werden.

---

Hierzu 6 Seite(n) Zeichnungen

---

FIG. 1

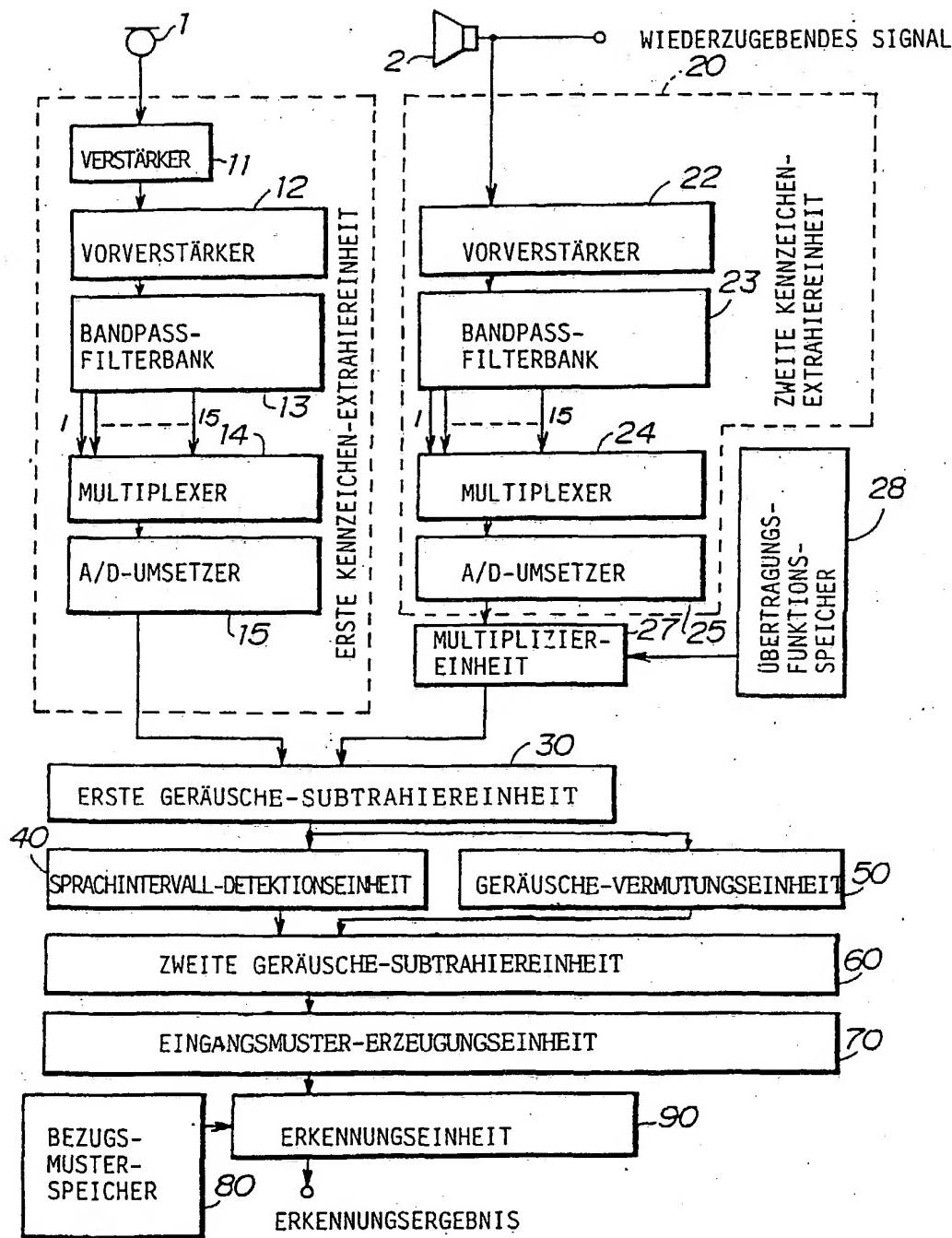


FIG. 2

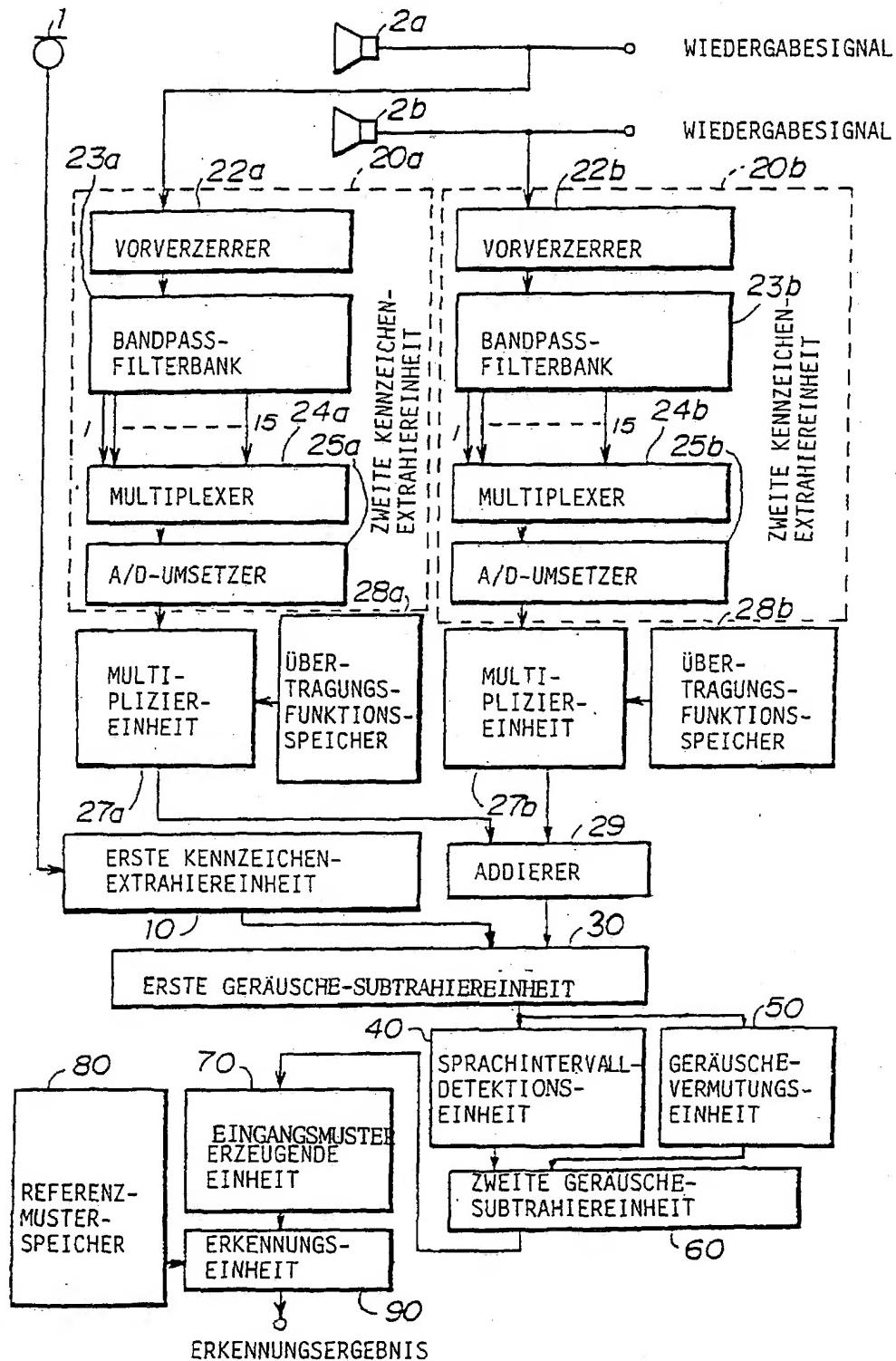


FIG. 3

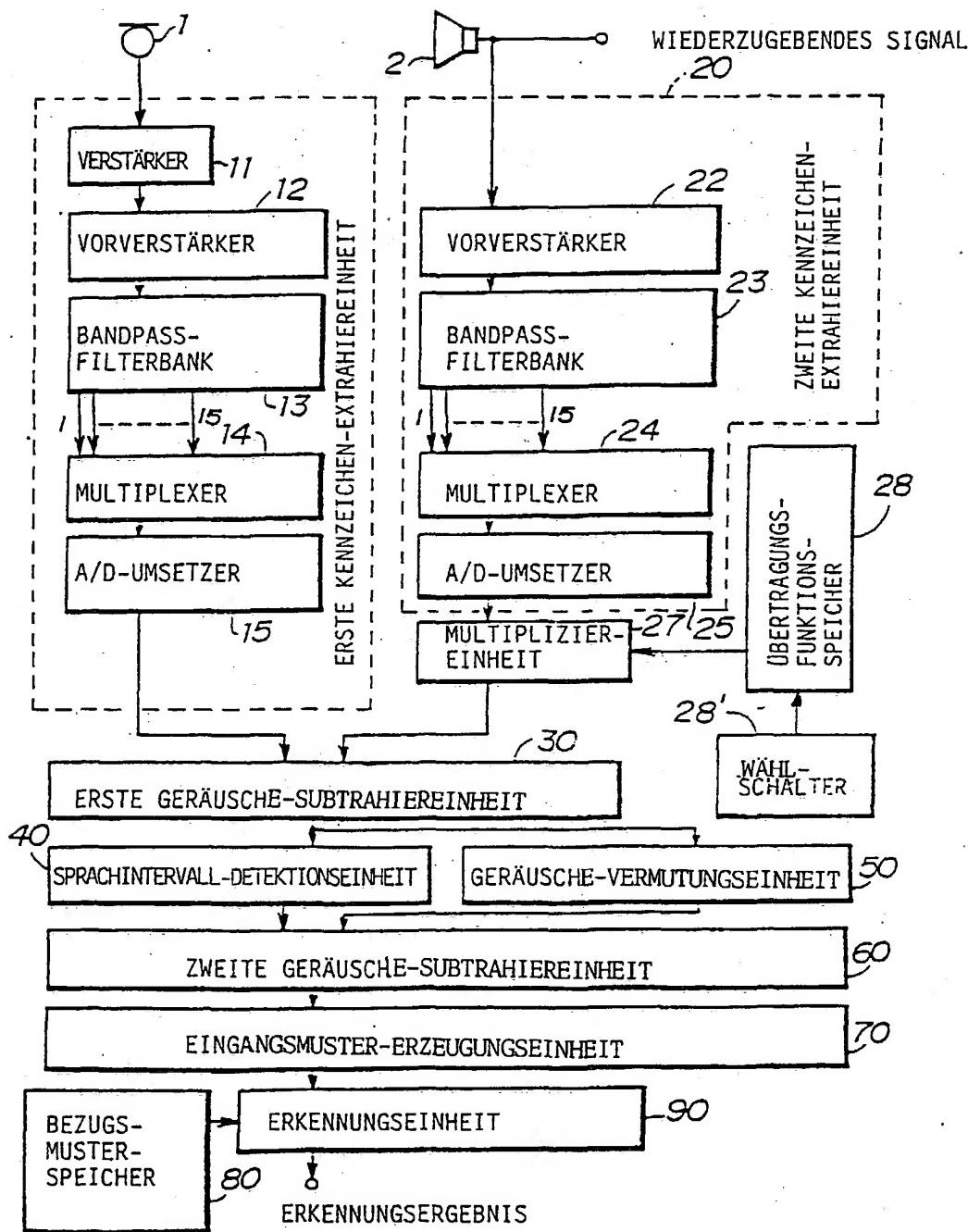


FIG. 4

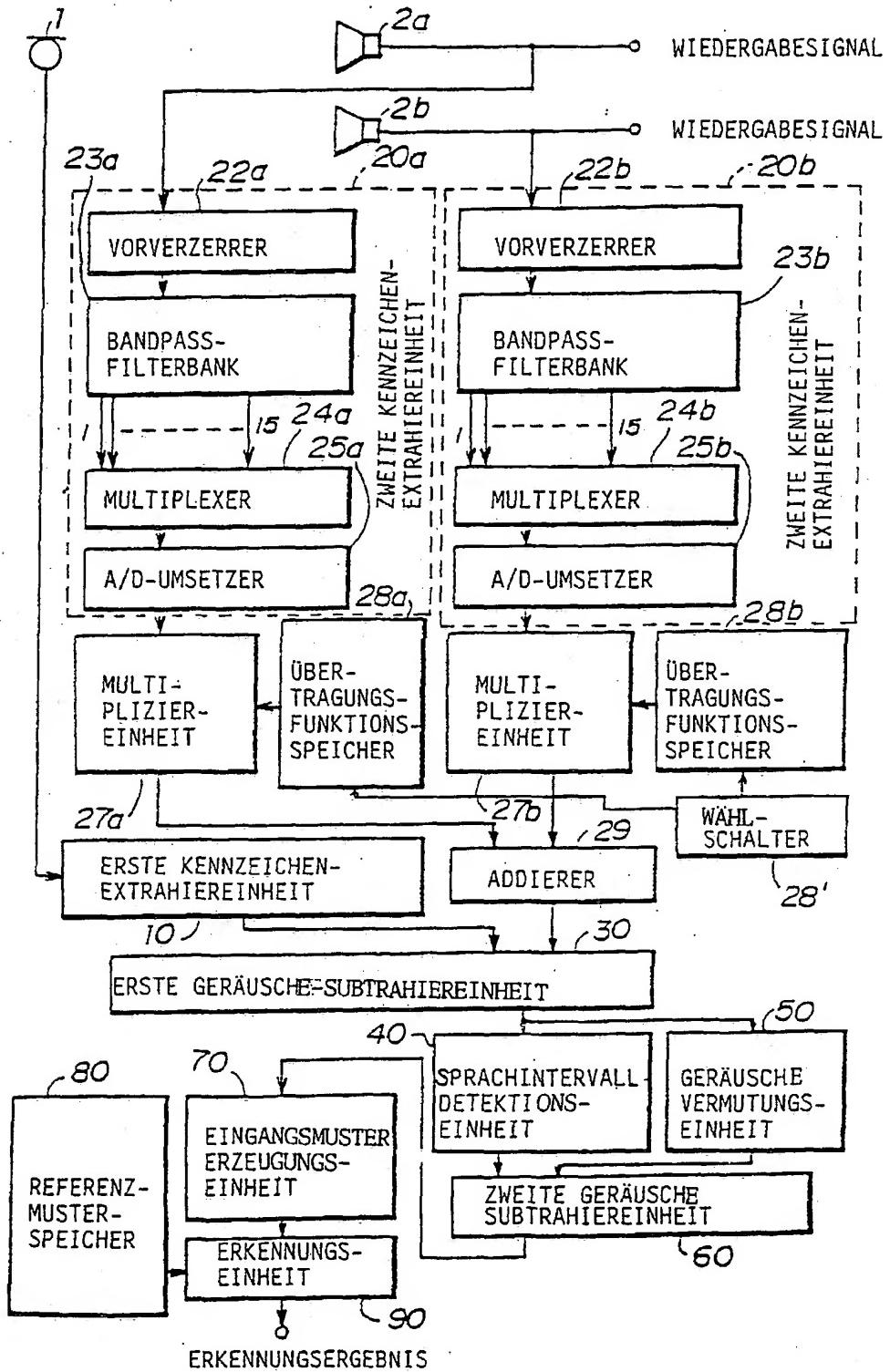


FIG. 5.

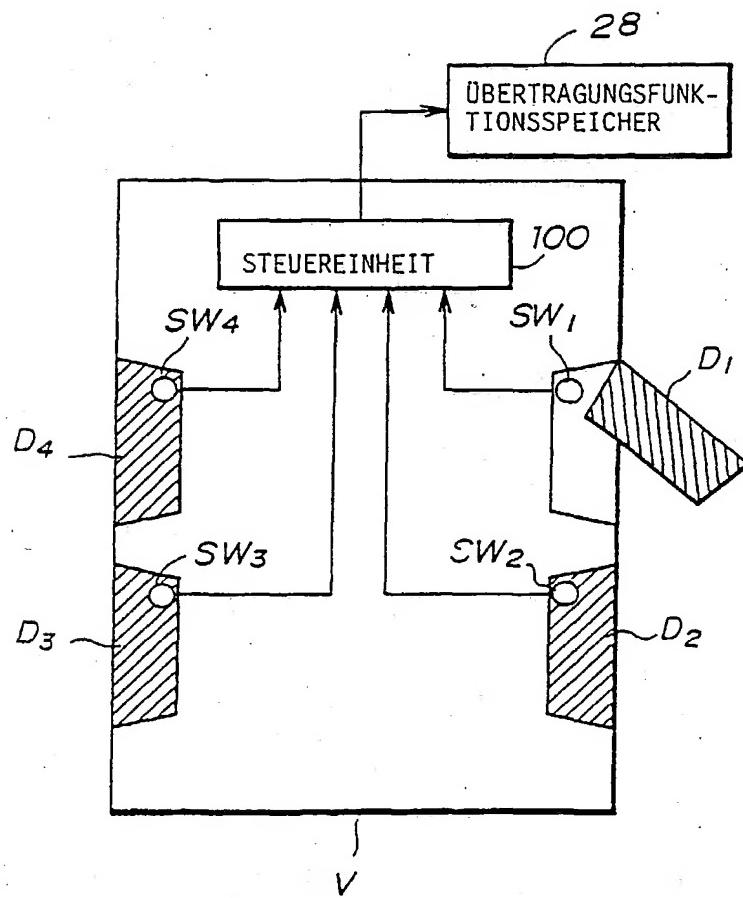


FIG. 6

